Does Rain Really Cause Toothache? Statistical Analysis Based on Google Trends

Se-Jeong Jeon[†]

Department of Dental Hygiene, Daejeon Institute of Science and Technology, Daejeon 35408, Korea

Background: Regardless of countries, the myth that rain makes the body ache has been worded in various forms, and a number of studies have been reported to investigate this. However, these studies, which depended on the patient's experience or memory, had obvious limitations. Google Trends is a big data analysis service based on search terms and viewing videos provided by Google LLC, and attempts to use it in various fields are continuing. In this study, we endeavored to introduce the 'value as a research tool' of the Google Trends, that has emerged along with technological advancements, through research on 'whether toothaches really occur frequently on rainy days'.

Methods: Keywords were selected as objectively as possible by applying web crawling and text mining techniques, and the keyword "b" meaning rain in Korean was added to verify the reliability of Google Trends data. The correlation was statistically analyzed using precipitation and temperature data provided by the Korea Meteorological Agency and daily search volume data provided by Google Trends.

Results: Keywords "chi-gwa", "chi-tong", and "chung-chi" were selected, which in Korean mean 'dental clinic', 'toothache', and 'tooth decay' respectively. A significant correlation was found between the amount of precipitation and the search volume of tooth decay. No correlation was found between precipitation and other keywords or other combinations. It was natural that a very significant correlation was found between the amount of precipitation, temperature, and the search volume of "b".

Conclusion: Rain seems to actually be a cause of toothache, and if objective keyword selection is premised, Google Trends is considered to be very useful as a research tool in the future.

Key Words: Google Trends, Meteorotropic toothache, Research tools

Introduction

Any Korean will be familiar with the expression 'It rains when grandma feels pain in her knees'. Also known as a 'meteorotropic disease', such absurd story is surprisingly widespread not only in Korea but also around the globe. In fact, numerous studies have been conducted on this condition. However, while a number of studies affirmed the relationship between weather or climate and chronic pain¹⁻⁵⁾, opposing researchers have classified this as a myth with insufficient scientific evidences⁶⁻⁸⁾.

The ultimate reason for the contrasting arguments regarding the same proposition is because the majority of

the studies relied on inaccurate and biased memories and experiences of humans, alongside the challenges to scientifically prove the hypothetical neurological and hormonal changes such as decreased serotonin secretion in the studies that affirmed the correlation between weather and pain.

It is possible for some researchers to propose studies based on statistics as a method to improve such studies. With the advent of the 4th industrial revolution and the big data industry exponentially emerging as a new hot item, there have been worldwide trends of revealing many data, which were once held by the public sectors, to the private sectors. The Republic of Korea is also supporting this

Received: April 30, 2021, Revised: May 24, 2021, Accepted: June 2, 2021

eISSN 2233-7679

[†]Correspondence to: Se-Jeong Jeon, <https://orcid.org/0000-0002-2005-8096>

Department of Dental Hygiene, Daejeon Institute of Science and Technology, 100, Hyecheon-ro, Seo-gu, Daejeon 35408, Korea
Tel: +82-42-580-6351, Fax: +82-42-580-6301, E-mail: sejeong0084@gmail.com

Copyright © The Korean Society of Dental Hygiene Science.

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

trend through the use of a public data portal (<https://data.go.kr/>). In the field of health care, the data necessary for research with no personal identification information can be downloaded through the open system on health care big data (<https://opendata.hira.or.kr/>) operated by the Health Insurance Review and Assessment Service, which can be an extremely reliable and useful tool as Korea imposes an obligation for all citizens to subscribe to the national health insurance.

However, the majority of the pain from a meteorotropic disease occur only for a short period during a specific weather condition before disappearing in many cases, and the degree of pain is generally mild. This prevents patients from readily making decisions to visit the hospital, and even when they do decide to visit, the pain perfectly vanishes by the time weather clears up, in which the patients decide to no longer visit the hospital. Even in the uncommon case of receiving treatment, the digital health insurance system only records the patient's date of visit, and it further becomes difficult to perform a statistical analysis because not all patients visit the hospital on the day of the pain onset.

Meanwhile, the popularization of smartphones has caused numerous individuals nowadays to promptly search their conditions online before visiting the hospital when they detect abnormal changes in their body, and the search results occasionally become more trusted than the opinions from experts. Google Trends is a free big data analysis service based on search terms and watched videos provided by Google (Google LLC, Mountain View, CA, USA), and became a matter of discussion when it accurately predicted Trump's overwhelming victory during the 2016 U.S. presidential election when the majority of polls were predicting the dominance of Hillary Clinton who was a Democratic candidate. Artificial intelligence (AI) and big data emerged as key keywords of the 4th industrial revolution following the incident, and attempts have been continuing to utilize the AI and big data in a variety of fields⁹⁾.

Although somewhat different depending on the querying conditions, Google Trends provides the daily search volume of keywords in a form that is very easy to process. Therefore, in this study, we examined the correlation

between weather and toothache, one of the dental symptoms of meteorotropic diseases, based on the Google Trends, and then tried to introduce the possibility of Google Trends as a new tool to researchers.

Materials and Methods

1. Limitation of scope and data collection

The weather data to be used as the basis of the statistics for this study were collected from the meteorological data open portal operated by the National Climate Data Center of the Korea Meteorological Administration (<https://data.kma.go.kr/>). There were slight concerns regarding restrictions of the weather data for Korea as Google is a search engine used worldwide, but the volume of search by foreigners was presumed to be minimal as only the keywords in Korean were to be selected for this study. The average daily temperature and daily precipitation in South Korea were collected, and unfortunately, humidity and air pressure data were not provided on a daily basis, so they could not be used for this study.

Search volume data was collected on the Google Trends web page (<https://trends.google.co.kr/>). In the past, the number of searches was provided as it is, but now, a relative number is provided with the highest search volume being 100 during the request period (Fig. 1). Data are provided in the form of graphs and csv files, and in this study, csv files were downloaded and processed to facilitate statistical analysis.

The investigation period was between January 1, 2020 to September 1, 2020. Although efforts were made to conduct the research on an annual basis as much as possible, limitations had to be unavoidably set as querying for data older than eight months only provided results in weeks or months and not as a daily unit.

Fig. 1. Example of Google Trends' provision of search volume.

2. Keywords selection

As with other studies based on unstructured data, keywords had to be chosen extremely carefully. If the researcher selects keywords arbitrarily, very biased research results can be derived in some cases, and the reliability of the research cannot be guaranteed. Therefore, existing literature that performed web crawling and text mining with a computer program was used for this study, to be as fair and objective as possible in selecting the keywords¹⁰.

Naver is the largest Internet portal site in South Korea, and after being launched in 1999, it is currently ranking first in the search market share. Jisik-iN is a knowledge exchange service between users provided by Naver, and it is a very large community with about 450 million accumulated answers since the service started in 2002. Naver supports open APIs for developers, and through this, it is possible to access the vast amount of data stored on their server more easily, quickly, and legally. Therefore, in this study, Jisik-iN question is collected based on the web crawling technique, the morpheme is analyzed by applying the natural language processing (NLP) technique included in the text mining technology, and finally keywords were selected based on the frequency of appearance of nouns.

1) Web crawling

A web crawling program was developed using Python,

```
import requests
import json
import openpyxl

keyword = '치통'
API_ID = 'q8tgZ4Nzi'
API_SECRET = '09'

url = "https://openapi.naver.com/v1/search/kin.json"
header = {
    'X-Naver-Client-Id': API_ID,
    'X-Naver-Client-Secret': API_SECRET
}
params = {'query': keyword, 'display': '100', 'sort': 'sim'}

wb = openpyxl.Workbook()
sheet = wb.active

for i in range(100):
    params['start'] = i+1
    r = requests.get(url, params=params, headers=header)
    j = json.loads(r.text)

    for k in j['items']:
        k['title'] = k['title'].replace("<br>", "").replace("</br>", "")
        k['description'] = k['description'].replace("<br>", "").replace("</br>", "")
        t = [k['title'], k['link'], k['description']]
        sheet.append(t)

wb.save('res.xlsx')
```

Fig. 2. Program source code based on Python produced for web crawling.

an open source-based programming language, and was designed to use the Open API provided by Naver (Fig. 2). Among the questions registered in the Jisik-iN service from Jan 1, 2019 to Dec 31, 2019, those including ‘*chi-tong*’, meaning toothache in Korean, were sorted by similarity and the top 10,000 were extracted (Fig. 3). The queries were only provided in partial according to Naver's policy, but there was a possibility of issues such as copyright infringement if web crawling was performed by other unauthorized methods without using the open API. In addition, the data provided in partial was considered ample to perform the analysis without concerns.

2) Text mining

KoNLPy is a open source based Python package for NLP of the Korean language, which enables the extraction of the necessary data by analyzing Korean sentences in units of morphemes¹¹. After extracting only the nouns from the aforementioned 10,000 collected queries using KoNLPy, the final keywords were selected based on the appearance frequency and several criteria.

3) Verification of reliability of Google Trends data

Since there has not yet been enough data accumulated to ensure the reliability of research using Google Trends, the proposition “People will search for rain more on rainy days” was assumed to be true and the correlation between daily search volume of “*bi*” meaning ‘rain’ in Korean and precipitation was investigated.

Fig. 3. Partial sector of the extracted Jisik-iN query from Naver.

3. Statistical analysis

The collected data was statistically analyzed using SPSS Statistics 20 (IBM Co., Armonk, NY, USA). A correlation analysis was performed between the daily

search volume of the selected keywords and the temperature and precipitation.

Results

1. Selected keywords

First, a total of 100 words were selected in the order of appearance frequency from the aforementioned 10,000 queries collected. Next, stopwords such as “*ttae-mun*”, “*gab-ja-gi*”, and “*jeong-do*” meaning ‘because of’, ‘suddenly’, and ‘degree’ each in Korean, and keywords that are difficult to adopt in this research such as “*oen-jjog*” (‘left’) or “*yang-jjog*” (‘both sides’) were removed. The words “*nei-beo*” (‘Naver’), “*jisik-in*” (‘Jisik-iN’), and “*hi-doc*” (‘doctors who donate talent in Jisik-iN’), which exist due to the nature of the data, were also excluded from the list. The keywords remaining so far were visualized in Fig. 4 using the Python wordcloud library.

After extracting the top 10 again from the remaining keywords, “*tong-jeung*” (‘pain’), “*chi-a*” (‘tooth’), “*chi-lyo*” (‘treatment’), “*i*” (‘tooth’), “*eo-geum-ni*” (‘molar’), “*du-tong*” (‘headache’), and “*sang-dam*” (‘consult’) which are

Fig. 4. Visual presentation of keyword candidates using the Python WordCloud library. Larger font sizes reflect higher frequency of appearance.

Table 1. Some Examples of Processed Data

Date	Daily precipitation (mm)	Average daily temperature (°C)	Keyword search volume (relative value)			
			“ <i>bi</i> ” (‘rain’)	“ <i>chi-tong</i> ” (‘toothache’)	“ <i>chi-gwa</i> ” (‘dental clinic’)	“ <i>chung-chi</i> ” (‘tooth decay’)
2020-01-01	0	−0.9	57	0	15	46
2020-01-02	0	1.6	73	35	57	20
2020-01-03	0	2.1	64	0	59	21
2020-01-04	0	1.6	62	0	11	0
2020-01-05	0	1.5	67	38	41	0
2020-01-06	10.0	3.5	63	34	42	40
...						
2020-05-06	0	17.8	55	26	27	15
2020-05-07	0	17.1	58	53	70	16
2020-05-08	3.7	17.2	60	0	60	99
2020-05-09	33.7	15.8	60	0	29	19
2020-05-10	0.4	15.9	70	0	25	0
2020-05-11	0.3	18.3	65	26	40	0
...						
2020-08-28	14.5	27.2	77	0	60	0
2020-08-29	12.9	26.8	73	0	29	18
2020-08-30	6.8	26.7	79	0	20	17
2020-08-31	2.7	25.7	73	0	38	29
2020-09-01	1.4	25.4	80	24	47	14

thought to be too general to produce meaningful results, were excluded, and finally, “*chi-tong*” (‘toothache’), “*chi-gwa*” (‘dental clinic’), and “*chung-chi*” (‘tooth decay’) were selected.

2. Data organization and processing

The collected data was organized and processed to facilitate statistical analysis. Table 1 shows some part of the finally processed data.

3. Statistical analysis

1) Correlation between weather data and search volume of “*bi*” (‘rain’)

There was a very significant correlation between daily precipitation and average daily temperature, and the amount of search for the keyword “*bi*” (‘rain’), at $p < 0.001$ (Table 2).

2) Correlation between weather data and search volume of selected keywords

A significant correlation was found between daily

precipitation and the amount of search for “*chung-chi*” (‘tooth decay’). There was also a significant correlation between the search volume for “*chi-gwa*” (‘dental clinic’) and the search volume for “*chung-chi*” (‘tooth decay’). There was no significant correlation between the other variables (Table 3).

Discussion

In 1992, Shetty et al.¹⁾ reported temperature and humidity to affect the degree of pain experienced by chronic pain patients, with the most common claims from patients being joint and muscle pain. In 2003, Sato²⁾ reported temperature and air pressure to increase the pain intensity in an animal experimental study. A plethora of other studies have also affirmed the correlation between weather and chronic pain³⁻⁵⁾. Conversely, in 1991, Clarke and Nicholl⁶⁾ stated weather to have no significant effect on pain, and Chami et al.⁷⁾ in 2006 stated that there were insufficient scientific evidences to the support the claims from the patients. In 1996, Redelmeier and Tversky⁸⁾ stated that the existing perceptions of patients play a

Table 2. Results of Correlation Analysis between the Weather Data and Keyword “*bi*” (‘rain’)

Variable	Daily precipitation (mm)	Average daily temperature (°C)	Keyword search volume (relative value)
			“ <i>bi</i> ” (‘rain’)
Daily precipitation (mm)	1		
Average daily temperature (°C)	0.308**	1	
Keyword search volume (relative value) “ <i>bi</i> ” (‘rain’)	0.327**	0.556**	1

All p-values by Pearson correlation. ** $p < 0.001$.

Table 3. Results of Correlation Analysis between the Weather Data and the Selected Keywords

Variable	Daily precipitation (mm)	Average daily temperature (°C)	Keyword search volume (relative value)		
			“ <i>chi-tong</i> ” (‘toothache’)	“ <i>chi-gwa</i> ” (‘dental clinic’)	“ <i>chung-chi</i> ” (‘tooth decay’)
Daily precipitation (mm)	1				
Average daily temperature (°C)	0.308**	1			
Keyword search volume (relative value)					
“ <i>chi-tong</i> ” (‘toothache’)	−0.066	−0.081	1		
“ <i>chi-gwa</i> ” (‘dental clinic’)	0.047	0.003	0.003	1	
“ <i>chung-chi</i> ” (‘tooth decay’)	0.157 (0.014)*	0.064	0.022	0.149 (0.020)*	1

All p-values by Pearson correlation. * $p < 0.05$, ** $p < 0.001$.

similar role to that of a placebo effect, which enables the patients to continue their unfounded beliefs.

Researchers of the past often relied on questionnaires to obtain the data they needed for research. For example, they threw questions like “Did your knee hurt on a rainy day?” or “Do you think there’s a connection between the weather and your pain?” and then said “Please choose between 1 and 10 points for how sick you were”. Therefore, somewhat subjective data were produced based on patient experience and memory. Of course, all researchers have been highly aware of such issues, and thereby established and attempted various devices to further objectify the data. However, structural limitations still exist nonetheless in a number of topics, as mentioned in the examples above. Therefore, this study intended to make an approach from a new standpoint.

Although studies which employed Google Trends have been gradually published, there is still an insufficient amount of accumulated data to guarantee its reliability. The algorithm used to calculate and provide the search volume data has not yet been disclosed by Google Trends, alongside the possibility of undetected errors. Beyond, the possibility of data tampering cannot be completely excluded due to the nature of a profit-seeking company. However, as we have seen in the previous US presidential election case, depending on the use, it may yield more valuable results than any other statistical data. Therefore, in this study, we first secured minimal reliability on Google Trends data by verifying the hypothesis that “People will search more for ‘rain’ on rainy days.”

Looking at the results of the study, a significant correlation was found between the daily precipitation and the amount of “*chung-chi*” (‘tooth decay’) search, in the end, it seems that toothache occurs more frequently on rainy days. Although the evidence presented in this study cannot be said to be superior to the previous studies, it has its own significance because it is overwhelmingly unlikely to correlate by chance, given the number of samples and significance level. There was no significant correlation between the amount of precipitation and the search volume of “*chi-tong*” (‘toothache’) or “*chi-gwa*” (‘dental clinic’), which seems to be because people think that the most common cause of toothache is tooth decay and

search it. However, more research is needed on this.

The topic selected for this study may appear to have miniscule academic significance at first glance. However, rather than focusing on verifying the hypothesis that ‘Toothache occurs more on a rainy day’, we put more significance in showing later researchers the ‘value as a research tool of the Google Trends’ that has emerged as a result of the development of civilization. As previously discussed, data from Google Trends can yield fairly accurate and intuitive results, and can be applied to a wide variety of studies depending on its use. If objective keyword selection is preceded, it is thought that very accurate and objective data can be obtained while saving effort, time, and cost.

Notes

Conflict of interest

No potential conflict of interest relevant to this article was reported.

Ethical approval

This study does not contain experiments that require ethical permission.

ORCID

Se-Jeong Jeon, <https://orcid.org/0000-0002-2005-8096>

References

1. Shetty Jr MS, Cundiff G, DeGood DE: Pain complaint and the weather: weather sensitivity and symptom complaints in chronic pain patients. *Pain* 49: 199-204, 1992.
[https://doi.org/10.1016/0304-3959\(92\)90143-y](https://doi.org/10.1016/0304-3959(92)90143-y)
2. Sato J: Weather change and pain: a behavioral animal study of the influences of simulated meteorological changes on chronic pain. *Int J Biometeorol* 47: 55-61, 2003.
<https://doi.org/10.1007/s00484-002-0156-9>
3. Aikman H: The association between arthritis and the weather. *Int J Biometeorol* 40: 192-199, 1997.
<https://doi.org/10.1007/s004840050041>
4. Guedj D, Weinberger A: Effect of weather conditions on rheumatic patients. *Ann Rheum Dis* 49: 158-159, 1990.

- <https://doi.org/10.1136/ard.49.3.158>
5. Ng J, Scott D, Taneja A, Gow P, Gosai A: Weather changes and pain in rheumatology patients. *APLAR J Rheumatol* 7: 204-206, 2004.
<https://doi.org/10.1111/j.1479-8077.2004.00099.x>
 6. Clarke AM, Nicholl J: Does the weather affect the osteoarthritic patient? *Br J Rheumatol* 30: 477, 1991.
<https://doi.org/10.1093/rheumatology/30.6.477>
 7. Chami G, Moulder E, Mohsen A: Weather associated symptoms after ankle injuries; myth or reality? *The Foot* 16: 165-168, 2006.
<https://doi.org/10.1016/j.foot.2005.12.003>
 8. Redelmeier DA, Tversky A: On the belief that arthritis pain is related to the weather. *Proc Natl Acad Sci U S A* 93: 2895-2896, 1996.
<https://doi.org/10.1073/pnas.93.7.2895>
 9. Jun SP, Yoo HS, Choi S: Ten years of research change using Google Trends: from the perspective of big data utilizations and applications. *Technol Forecast Soc Change* 130: 69-87, 2018.
<https://doi.org/10.1016/j.techfore.2017.11.009>
 10. Kim BR, Ahn E, Hwang SJ, Jeong SJ, Kim SM, Han JH: Analysis of dental hygienist job recognition using text mining. *J Dent Hyg Sci* 21: 70-78, 2021.
<https://doi.org/10.17135/jdhs.2021.21.1.70>
 11. Park EJ, Cho SZ: KoNLPy: Korean natural language processing in Python. In: sigHCLT, ed. *Proceedings of the 26th Annual Conference on Human and Cognitive Language Technology*. sigHCLT, pp.133-136, 2014.